

# Using machine learning for transient classification in searches for gravitational-wave counterparts

Cosmin Stachie,<sup>1★</sup> Michael W. Coughlin<sup>2,3</sup>, Nelson Christensen<sup>1</sup>  
and Daniel Muthukrishna<sup>4</sup>

<sup>1</sup>*Artemis, Université Côte d'Azur, Observatoire Côte d'Azur, CNRS, CS 34229, F-06304 Nice Cedex 4, France*

<sup>2</sup>*Division of Physics, Math, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>3</sup>*School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA*

<sup>4</sup>*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2020 June 13. Received 2020 June 12; in original form 2019 December 12

## ABSTRACT

The large sky localization regions offered by the gravitational-wave interferometers require efficient follow-up of the many counterpart candidates identified by the wide field-of-view telescopes. Given the restricted telescope time, the creation of prioritized lists of the many identified candidates becomes mandatory. Towards this end, we use *astorapid*, a multiband photometric light-curve classifier, to differentiate between kilonovae, supernovae, and other possible transients. We demonstrate our method on the photometric observations of real events. In addition, the classification performance is tested on simulated light curves, both ideally and realistically sampled. We show that after only a few days of observations of an astronomical object, it is possible to rule out candidates as supernovae and other known transients.

**Key words:** gravitational waves.

## 1 INTRODUCTION

The first detection of a binary neutron star (BNS) system GW170817 (Abbott et al. 2017a) by the gravitational-wave (GW) detectors Advanced LIGO and Advanced Virgo was accompanied by the detection of both a short gamma-ray burst (SGRB) by *Fermi* Gamma-Ray Burst Monitor (GBM) (Abbott et al. 2017c; Goldstein et al. 2017; Savchenko et al. 2017) and a kilonova by many other facilities (Abbott et al. 2017d; Coulter et al. 2017; Smartt et al. 2017). This *kilonova* is the ultraviolet/optical/infrared emission powered by the neutron-rich outflows undergoing the radioactive decay of r-process elements (Lattimer & Schramm 1974; Li & Paczynski 1998; Metzger et al. 2010; Kasen et al. 2017). The specifics of the light curves of kilonovae depend on the equation of state of neutron stars and the mass ratio of the binary (Bauswein, Baumgarte & Janka 2013; Piran, Nakar & Rosswog 2013; Abbott et al. 2017a; Bauswein et al. 2017; Dietrich & Ujevic 2017; Radice et al. 2018). In addition to this, there is synchrotron emission, which arises from a compact central engine launching a highly relativistic jet of electron/positron/baryon plasma (Wijers, Rees & Mészáros 1997; Mészáros & Rees 1998). The internal dissipation of the jet's energy is responsible for the production of gamma-rays and hard X-rays. The *afterglow* phase, produced by interaction of the jet with the ambient material, consists of long-lasting multiwavelength emission in the X-ray, optical, and radio. These three possible electromagnetic signatures of GW events, the kilonova, the SGRB, and the afterglow, have different characteristics. The kilonova is a short-lived isotropic emission in

the visible and near-infrared spectrum, the SGRB is a beamed flare of high-energy X-rays and gamma-rays having a duration lower than 2 s, and the afterglow is a long-standing multiwavelength transient. There have been a number of examples in the literature of using the photometry of both afterglows (Troja et al. 2018; Ascenzi et al. 2019) and kilonovae (Coughlin et al. 2017; Smartt et al. 2017; Coughlin et al. 2018a) to place constraints on the character of the progenitor systems.

The joint observations of these systems are interesting for a variety of reasons, including the study of SGRB beaming, energetics, and galactic environment (Metzger & Berger 2012). In addition, the study of the kilonova light curves provides precious information about the nucleosynthesis of heavy elements in the Universe (Drout et al. 2017; Kasen et al. 2017; Pian et al. 2017; Watson et al. 2019) and the Hubble constant (Abbott et al. 2017b; Coughlin et al. 2019a). The first SGRB detected in association with a kilonova was 130603B (Tanvir et al. 2013), providing support for the existence of an unbeamed electromagnetic signature to compact binary mergers. Also, GRB 150101B has been reported as an off-axis jet associated with a blue kilonova by Troja et al. (2018), based on its resemblance to GRB 170817A. However, the detection of the kilonova transient represents a difficult task given the large sky localizations provided by both the gamma-ray satellites, such as the *Fermi* GBM and GW interferometers. The localizations released by the GW detectors, in particular, can be large, spanning  $\approx 100 - 10\,000$  deg<sup>2</sup> (Röver et al. 2007; Fairhurst 2009; Wen & Chen 2010; Fairhurst 2011; Grover et al. 2014; Sidery et al. 2014; Singer et al. 2014; Berry et al. 2015; Cornish & Littenberg 2015; Essick et al. 2015; Klimentko et al. 2016). While GRB detections only have 2D sky localization information, the strain measurement in GW events also allows for the computation of

★ E-mail: scosmin@oca.eu

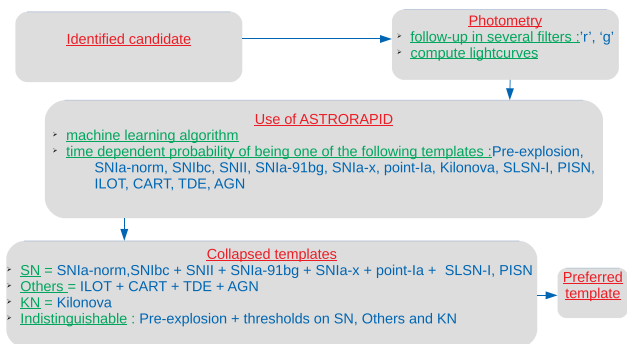
a luminosity distance and therefore complete 3D skymap information is provided.

The large sky localization regions require the use of wide-field survey telescopes to be covered. Observing instruments such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Morgan et al. 2012), Asteroid Terrestrial-impact Last Alert System (ATLAS; Tonry et al. 2018), the Zwicky Transient Facility (ZTF; Bellm et al. 2018; Graham et al. 2019), and in the near future BlackGEM (Bloemen et al. 2016) and the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2019) have the capabilities to observe these sky localizations. However, the main difficulty remains in the large number of contaminant transients detected each night by these surveys. Among them, the identification of the counterpart represents a significant challenge. For this reason, an effective follow-up requires coordination between the wide FOV telescopes discovering transients and the telescopes that perform the follow-up and characterization of those transients. The transient characterization is typically done by smaller FOV telescopes performing both photometry and spectroscopy. Since the available telescope time at these telescopes is limited, it is essential to minimize the number of candidates that require observations, and be as efficient as possible with the classification.

The challenging follow-up of electromagnetic transients has pushed the astronomical community to find new ways to optimize searches. An example is the formation of telescope networks (Coughlin et al. 2019d; Antier et al. 2020) that are generally built around the previously mentioned synoptic systems. Improved ways of tiling and observing these localizations using allocated time of telescopes are being employed based on the GW trigger candidate, telescope configuration, and possible electromagnetic counterparts (Coughlin et al. 2018b). New proposals for the amelioration of the galaxy targeting strategy have been suggested, such as the prioritization of galaxies based on their stellar mass (Ducoin et al. 2020).

Techniques to optimize the follow-up of objects have been proposed in the literature. In Coughlin et al. (2019d), a method that combines the automated filtering and human vetting is used in order to reduce the number of initial candidates for S190425z (LIGO Scientific Collaboration & Virgo Collaboration 2019), the first BNS candidate from the third Advanced LIGO–Advanced Virgo observing run. During the automatic analysis, asteroids or near-Earth objects are removed as they did not appear in consecutive observations separated by a few tens of minutes. Objects very close ( $<2$  arcsec) to point-like sources or having a historical detection prior to three days before the trigger are also automatically rejected. Finally, machine learning algorithms are used to identify image artefacts. Altogether, due to this automated filtering, it was possible to reduce the number of candidates from more than 300 000 to less than 300. Then, human vetting kept only those triggers that are in the localization, both in the 2D and distance, and that exhibited a rapid colour evolution consistent with a kilonova. At the end of the entire analysis, fewer than 20 candidates remained. As another example, in Andreoni et al. (2020), one can see the importance of having data not only after the GW trigger, but also prior to it. The DECAM follow-up of the GW alert S190814bvs showed that it was difficult to rule out candidates due to the lack of recent pre-imaging history, resulting in significantly more candidates despite a smaller localization to cover. A proposal for improving the training set destined to a machine learning algorithm whose purpose is to identify supernova photometric light curves is in Ishida et al. (2019).

There are two kinds of objects that exhibit time variability: astrophysical objects whose signal lasts for a limited time and



**Figure 1.** Flow chart illustrating the different steps made in the photometry analysis. The starting point is represented by the initial set of identified candidates, after which optical observations are carried out for each of these events. Then *astorapid* provides a time-dependent probability distribution spread over 14 possible candidate classes. The results provided by *astorapid* are handled in order to give weights and discriminate between only four main classes: ‘SN’, ‘KN’, ‘Others’, and ‘Indistinguishable’. Finally, a preferred class is declared.

objects with periodic variation of the flux. Our knowledge about these electromagnetic events has dramatically improved in the last years. For some types, such as supernovae, there are models for their associated light curves, estimates of the occurrence rate, and studies of their host-galaxy environment. Because of this, it is possible to simulate what a specific telescope will observe taking into account the instrument sensitivity and sky background. This was performed, for example, in the LSST *PLATiCC* data challenge (Kessler et al. 2019). In this study, they considered both extragalactic and galactic transients. Then, by means of the SuperNova ANALYSIS software (SNANA; Kessler et al. 2009), a realistic set of light curves is generated illustrating what would be the LSST detections of transients of this type in the coming years. In this paper, we will use *astorapid* (Muthukrishna et al. 2019), a classifier tool based on machine learning to classify objects. It was trained on a set of light curves generated using SNANA and *PLATiCC* in order to simulate a realistic set of events that would be observed by ZTF. *astorapid* is designed to distinguish between transient templates.

We will evaluate the ability to use machine learning classifiers on early photometric light curves to support prioritization of transients for follow-up in GW and SGRB follow-up. While we will focus on kilonovae, the technique will be suitable for detection of afterglows. We will describe the algorithm we use in Section 2. In Section 3, we describe the performance of the algorithms. In Section 4, we offer concluding remarks and suggest directions for future research. The paper ends with Appendix A presenting the statistical justification for some classification criteria introduced in Section 2.

## 2 ALGORITHM

The idea of this analysis is to use multi-epoch photometry to identify interesting candidates. For the kilonova models being explored here, significant changes in magnitude are expected on time-scales of a single night. For this reason, telescope network photometry will determine which transients can feasibly be related to the event, and otherwise determine the background supernovae and other unrelated transients. A flow chart showing up the method used in this study (and explained further in the following) to identify and characterize optical counterparts is shown in Fig. 1.

For this purpose, we use *astorapid*, which was developed to distinguish between 14 different templates: ‘Pre-explosion’ [a template introduced in order to distinguish the targeted flaring event from the moments preceding it (Muthukrishna et al. 2019)], ‘SNIa-norm’ (a subtype of type Ia supernovae), ‘SNIbc’ (a subtype of core-collapse supernovae), ‘SNII’ (a subtype of core-collapse supernovae), ‘SNIa-91bg’ (a subtype of type Ia supernovae), ‘SNIa-x’ [a subtype of type Ia supernovae (Silverman et al. 2012; Foley et al. 2013)], ‘point-Ia’ [hypothetical supernova type (Shen et al. 2010)], ‘Kilonova’ [electromagnetic counterpart to either BNS or neutron star–black hole (NS–BH) mergers (Abbott et al. 2017a)], ‘SLSN-I’ [type I superluminous supernovae (Quimby et al. 2008)], ‘PISN’ [pair-instability supernovae (Ren, Christlieb & Zhao 2012)], ‘ILOT’ [intermediate-luminosity optical transients (Berger et al. 2009)], ‘CART’ [calcium-rich gap transients (Lunnan et al. 2017)], ‘TDE’ [tidal disruption events (Rees 1988)], and ‘AGN’ (active galactic nucleus). A description of each specific template can be found in Kessler et al. (2019). At the same time, the model libraries used in this analysis are publicly available.<sup>1</sup> Muthukrishna et al. (2019) describe the intrinsic luminosity, light-curve shapes, and colour evolution for the transient templates used in the training of *astorapid*. In addition, fig. 2 of Muthukrishna et al. (2019) displays example light-curve shapes of these templates. The tool takes as input data the light curve of the transient, including the time of the exposure, apparent magnitude, and associated error bar; the output consists of a time-dependent discrete probability distribution. It can take either a redshift (such as from a probable host galaxy) or not, and we will show results for both cases in the following. This distribution provides the probability for each one of the 14 template types. The probability distribution changes with every new observation, and the more points the light curve contains, the more precise the identification.

While *astorapid* was designed to classify full light curves, the goal of this analysis is to determine, given a few observations, how to prioritize objects for follow-up to support kilonova identification. To this end, we made a few modifications to the initial code. First of all, we collapse all 14 different templates into three main classes. Thus, we consider the following classes: ‘SN’ (which accounts for ‘SNIa-norm’, ‘SNIbc’, ‘SNII’, ‘SNIa-91bg’, ‘SNIa-x’, ‘point-Ia’, ‘SLSN-I’, and ‘PISN’), ‘Others’ (accounting for ‘ILOT’, ‘CART’, ‘TDE’, and ‘AGN’), and ‘KN’ (which is simply the ‘Kilonova’ template). The probability of ‘Others’ and ‘KN’ is simply the sum of the probabilities of their constituents. For the new probability for the ‘SN’ class in particular, we found that we are required to penalize it more than the others (likely because it is made up of a majority of the classes). Therefore, for ‘SN’ in particular, we use the sum of the probabilities of its components multiplied by the factor  $(1 - e^{-k_{\text{obs}}^{\text{th}}/\beta})$ , where  $k_{\text{obs}}^{\text{th}}$  stands for the  $(k + 1)$ th observation. The choice of the factor  $(1 - e^{-k_{\text{obs}}^{\text{th}}/\beta})$ , is based on the following considerations, and specific tests done with real data of transients detected by ZTF as described in Appendix A. Considering only the very first observations the probabilities of the initial *astorapid* templates are about equal, and thus the probability of the ‘SN’ class becomes large because it accounts for numerous initial templates, and needs to be reduced. We choose an exponential function to impose our knowledge that having only a few observations is uninformative (i.e. returning essentially the prior) and to leave unchanged the late observation predictions. Hand-tuning for  $\beta$  led to  $\beta = 4$ , although it is worth acknowledging that this particular choice might not be optimal.

Then, we also introduce a new class called ‘Indistinguishable’, where we will say that the preferred event is ‘Indistinguishable’ if none of the other classes (‘SN’, ‘KN’, and ‘Others’) has a probability higher than 40%. The 40 % threshold was selected at the end of several trials. This value represents a trade-off between two different behaviours. A higher threshold will favour too much the ‘Indistinguishable’ class at least for the very first observations where there is not much information and the initial weight of the ‘Pre-explosion’ template is already high. On the other hand, the consequence of a lower threshold will be to force the modified classifier to choose some class in {‘SN’, ‘KN’, ‘Others’}, although there is not enough information for any inference; this is also an undesirable effect. In summary, in total there are four classes: ‘KN’, ‘SN’, ‘Others’, and ‘Indistinguishable’. We also now introduce the idea of a ‘preferred event’ after  $k$  observations. We define the preferred event as  $X$  (here  $X$  stands for ‘KN’, ‘SN’, or ‘Others’) if two conditions are fulfilled: (i)  $\text{Prob}(X) = \max[\text{Prob}(\text{‘SN’}), \text{Prob}(\text{‘KN’}), \text{Prob}(\text{‘Others’})]$  and (ii)  $\text{Prob}(X) > 40\%$ . This will be convenient for classification later. More details about the motivation of both the penalty factor and the threshold can be found in Appendix A.

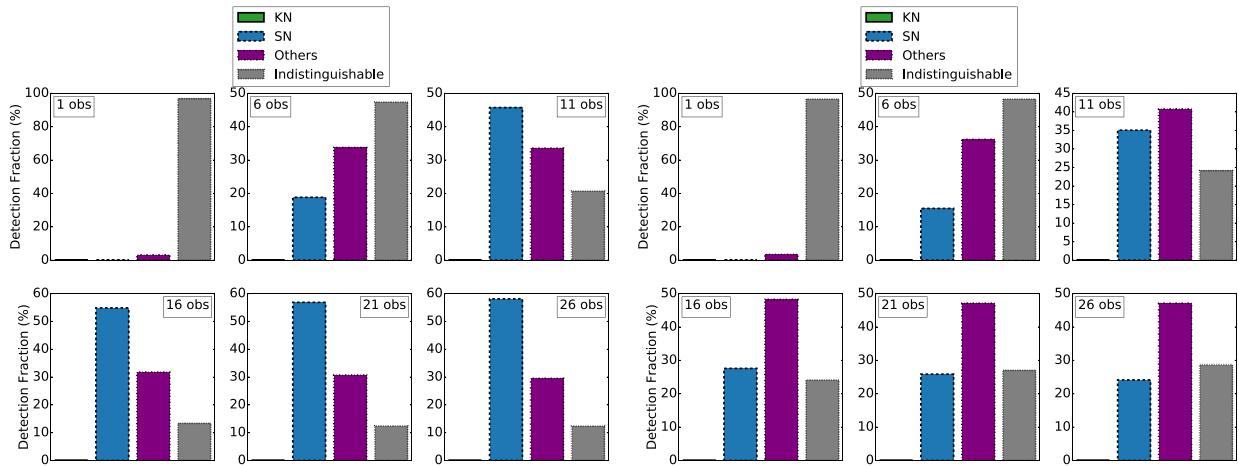
### 3 PERFORMANCE

To demonstrate the utility of the method for transient prioritization and identification, we seek to show that two conditions hold. The first is that kilonova light curves should in general be identified in the ‘KN’ class. The second is that for input light curves representing some other transient type, the analysis should not misidentify them as a ‘KN’. In the following, we will use both simulated light curves and real ZTF astrophysical transients from the public survey to assess these questions. For the injection sets in particular, we will create two sets of simulated SN and KN light curves. Each set has 1000 light curves representing transients uniformly distributed in distance between 40 and 3000 Mpc. The set of real ZTF events is formed by 2291 light curves from the public data stream with an average of 29 observations per light curve. These events are mainly different types of supernovae, but also include transients like TDE, AGN, ILOT, and CART.

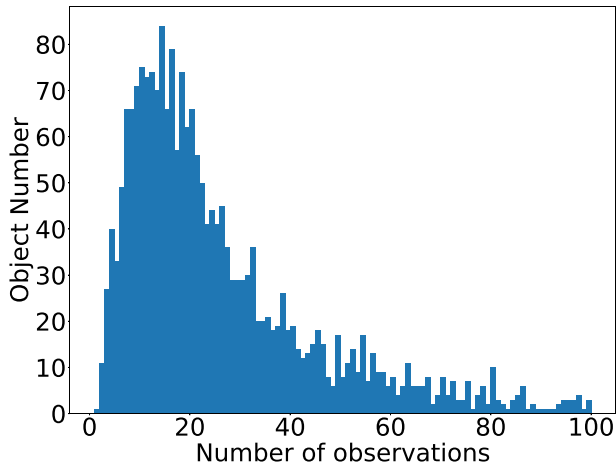
#### 3.1 Real transients observed in multiple filters

The study of light curves observed during real survey is obviously necessary to assess the performances of our classifier in realistic situations. To this end, we use public ZTF light curves. We put these events in two categories: ‘SN’ and ‘Others’. There are 2049 ‘SN’-type events (1450 ‘SNIa’, 110 ‘SNIbc’, 447 ‘SNII’, and 42 ‘SLSN’) and 174 ‘Others’-type events (152 ‘AGN’, 4 ‘CART’, 6 ‘ILOT’, and 12 ‘TDE’). As input, we consider only the observational data from the  $r$  and  $g$  bands. Fig. 2 displays the results of the classifier after the first observation points in the case of ‘SN’- and ‘Others’-type real events. One can see that the classifier starts to correctly classify (efficiency higher than 40%) after only 11 observations in the identification of ‘SN’ and 16 in the ‘Others’ case. It also exhibits that the classifier almost never misidentifies these real events as being kilonovae. This shows that kilonovae represent a fundamentally different part of the parameter space. As pointed out above, these real ZTF objects have an average of 29 observations per light curve. Fig. 3 illustrates the number of observations for the set of real ZTF objects. To take into account the unequal number of observations, we consider for a light curve possessing a total of  $N$  observations that the preferred event after  $m$  observations with  $m > N$  is the same as the one found at the end of the  $N$  (real) observations.

<sup>1</sup><https://doi.org/10.5281/zenodo.2612896>



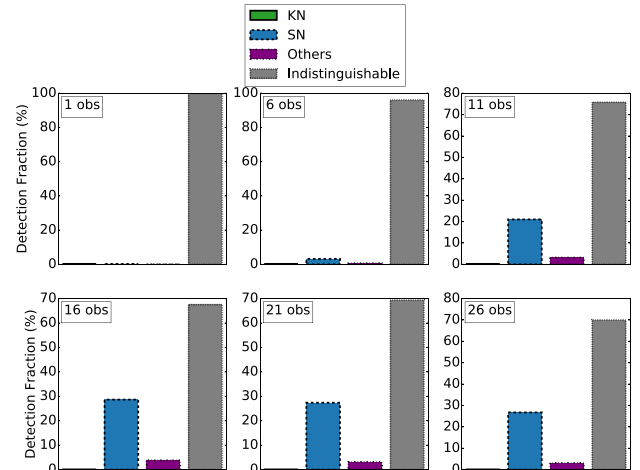
**Figure 2.** On the left are the histograms of classifier favoured events given different numbers of observations: from 1 (top-left corner) up to 26 (bottom-right corner). The input is represented by the set of 2049 ZTF real sources identified as ‘SN’ type. On the right is the same for the set of 174 ZTF real sources identified as the ‘Others’ type.



**Figure 3.** Histogram showing the observation number for the set of real ZTF objects. Both ‘SN’- and ‘Others’-types objects are included in this study.

### 3.2 Real transients observed in a single filter

Among the real supernova transient light curves, there are 157 transients that have only single passband observations. As for any deep learning algorithm, the idea is that the more information there is, the better is the classification. Thus, a classification result comparison between the two passbands ( $r$  and/or  $g$ ) case and only one passband (only  $r$  or only  $g$ ) case is worthwhile. In Fig. 4, we show the results of the follow-up of those single filter events. In this case, the success rate after 26 observations is much lower (around 10%) compared to the case presented in the previous section (success rate around 60%). This highlights how essential colour information is in the classification. It is worth mentioning that simulations for which *astrorapid* was trained on did not include single filter light curves; therefore, as a future activity retraining *astrorapid* to include both single and multiphotometric bands will help to fix this issue.



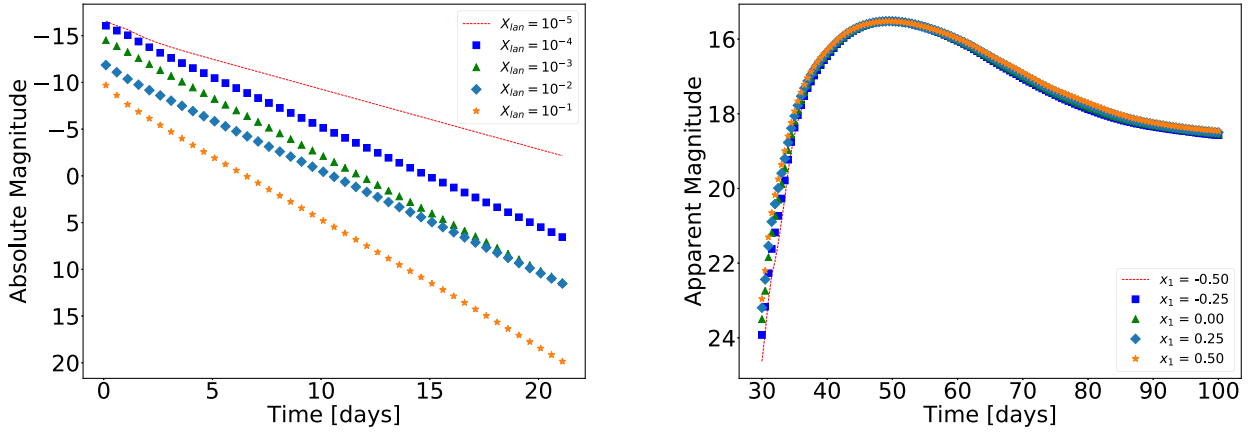
**Figure 4.** Preferred template fraction given different numbers of observations: from 1 (top-left corner) up to 26 (bottom-right corner). The input is represented by the set of 157 ZTF real sources observed in only one filter. Those events were finally all identified as being SN types.

### 3.3 Injection sets

There are several reasons that motivate us to verify performance using simulated light curves. It was shown in the previous sections how well the classifier performs in identifying supernovae sampled at the ZTF cadence (an illustration of the ZTF cadence is available in Fig. 14). A question that arises is how much the background transient identification can be improved if we increase the observational rate, or equivalently, how much a dedicated ToO observation will ameliorate the results compared to the typical ZTF cadence. Likewise, we want to check how well the kilonova light curves are recovered by our classifier. Because of the very small set of real kilonovae detected to date, the choice of an injection set becomes important.

As described above, we use simulations of both kilonova and supernova light curves generated by varying different parameters. The codebase used to generate kilonova light curves was described previously in Coughlin et al. (2018a, 2019b). The variable physical



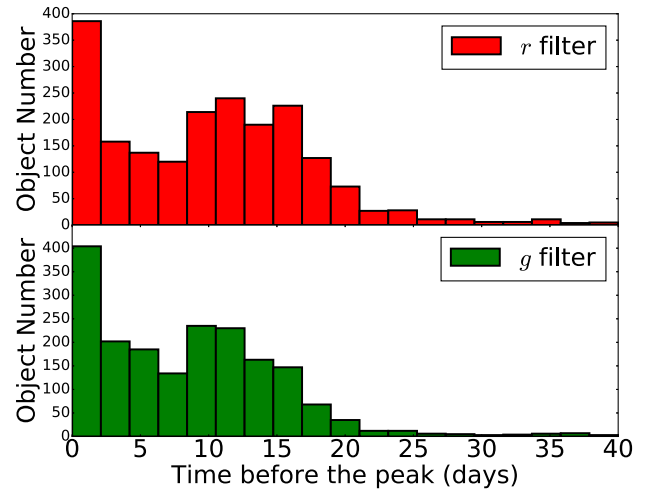


**Figure 5.** Left: Light curves representing the absolute magnitude as a function of time for five KN injections, for different input parameters. For all the light curves  $M_{\text{ej}} = 0.05 M_{\odot}$  and  $v_{\text{ej}} = 0.15c$ . The only parameter that varies is  $X_{\text{lan}}$ , whose value is shown on the legend. Right: Apparent magnitude versus time for five SN injections. The light curves share parameters like cosmological redshift  $z = 0.022$  and colour index  $c = 0$ . Here the shape parameter  $x_1$  is different for each light curve and its value is shown in the legend. These are the simulation results for the  $g$  band for both plots.

parameters in the model are the ejectum mass ( $M_{\text{ej}}$ ), the velocity of the ejecta ( $v_{\text{ej}}$ ), and the lanthanide fraction ( $X_{\text{lan}}$ ). Likewise, the prior ranges of the parameters are  $M_{\text{ej}} \in [0.01M_{\odot}, 0.1M_{\odot}]$ ,  $v_{\text{ej}} \in [0.01c, 0.3c]$ , and  $X_{\text{lan}} \in [10^{-5}, 10^{-1}]$ , where  $M_{\odot}$  is the solar mass and  $c$  represents the speed of light. While  $v_{\text{ej}}$  is sampled uniformly, the parameters  $X_{\text{lan}}$  and  $M_{\text{ej}}$  are log-uniformly distributed. An illustration of several such light curves is given on the right of Fig. 5. Type Ia supernova light curves are generated by ‘sncosmo’, whose details are explained in Guy et al. (2007) and Barbary et al. (2016). In order to be as general as possible, parameters like parameter shape (hereafter  $x_1$ ) and colour index (hereafter  $c$ ) were chosen to fill a broad space:  $x_1 \in [-0.5, 0.5]$  and  $c \in [-0.05, 0.05]$ . The left of Fig. 5 shows a few such light curves.

We want to imitate the ‘background’ introduced by such light curves in searches for kilonovae. To this end, it is important to understand where the supernova observations are positioned relative to the peak of the light curve. If there are enough points before the peak, then one could expect a few misclassifications as ‘KNe’ because a rising light curve is not characteristic for kilonovae. In Fig. 6, we illustrate the time of the first observation with respect to the peak of the supernova light curve. It is worth mentioning that by first observation, we mean first detection of the supernova. The data set used in this study was the same set of real ‘SN’ objects used in Section 3.1. It is worth mentioning that we cannot be sure if the peak of the light curve coincides with the peak of the supernova, while in principle, it could be a local maximum. However, even a local maximum will appear as a ‘rise’, which is one of the main features we are investigating here. From Fig. 6, one sees that, for a non-negligible (more than 10%) part of the objects, their first observation represents the peak of the light curve. We do not know if it coincides with the global peak or where the global peak might actually be. According to these results, we choose the first observations of the supernovae to be uniformly distributed in the range  $[-7 \text{ days}, +30 \text{ days}]$ .

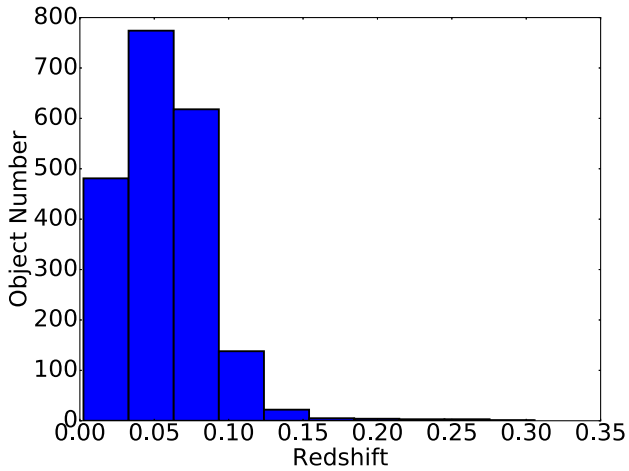
Looking at the distance distribution, we considered a uniformly distributed population of supernovae and kilonovae between 40 and 300 Mpc (chosen as the edge of the BNS GW detection horizons). This choice is also motivated by the distribution of the cosmological redshifts in the case real ZTF ‘SN’ objects; this is the same light curve set used in Section 3.1. Indeed in Fig. 7, one can see that this distribution possesses a peak around  $z = 0.05$ , which corresponds to



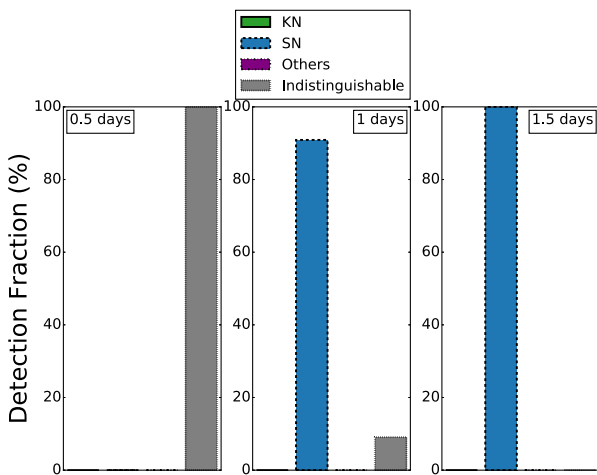
**Figure 6.** The amount of time before the peak of the light curve for the first observation. On the top (bottom) plot, there is the histogram corresponding to the  $r$  ( $g$ ) passband. The input is represented by the set of real ‘SN’ ZTF objects.

a luminosity distance of approximately 230 Mpc. Although a population uniformly distributed in volume is also possible, we injected uniformly in distance because we are predominantly interested in how the algorithm performs as a function of distance, and so how it performs spanning the bright to the faint end. A volume-limited set would be a more realistic distribution of kilonovae, and essential for any rates-related work.

In the case of kilonovae, we considered two sets: the first one ideally sampled (two observations per night per filter on average) and the second one realistically sampled (one observation per three nights per filter on average). In the view of already treated real ZTF ‘SN’ objects, we limit ourselves to only the case of ideally sampled supernova injections. Magnitude uncertainties have been also taken into account for these sets of light curves. The error bars considered for this study are magnitude and filter dependent, and have similar values to those measured on the real ZTF objects.



**Figure 7.** Redshift histogram for the real ZTF ‘SN’-type light curves.



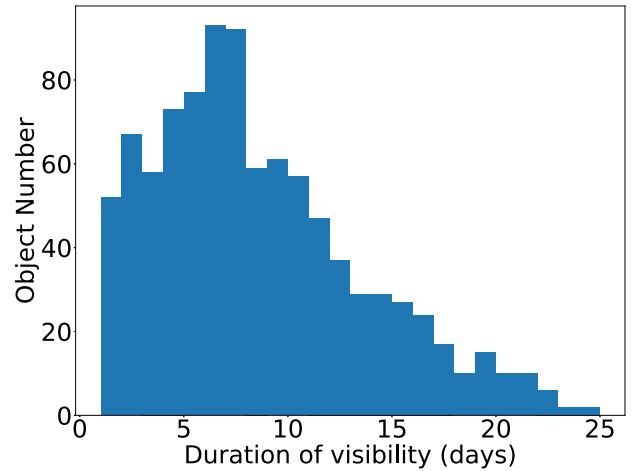
**Figure 8.** Preferred event fraction histogram given different amounts of observation time: from 0.5 day (left) up to 1.5 days (right). The input is represented by the set of 1000 SN injections.

We verify the performance of our method on the simulated light curves. In the case of kilonova injections, the output assesses how well the classifier recognizes a kilonova; the purpose of evaluating the output on supernova injections is to quantify how often our tool misidentifies a transient as being a kilonova when it is not. In this regard, the supernova transients represent our ‘background’. The choice of supernovae as the background set is motivated by the dominance of this type of transient among the identified candidates.

### 3.3.1 Well-sampled type Ia supernovae

In this section, the light curves are ideally sampled (two observations per night taken in each consecutive night) with detections in two filters ( $r$  and  $g$  bands). In the case of the type Ia supernova injections (see Fig. 8), the ‘SN’ template is preferred after only a few nights. This is because these are the brightest transient events, and therefore their brightness is the key determinant.

Also, given that these injections present only-rising, only-decreasing, and rising-decreasing shapes, one can conclude that the classification efficiency is not simply due to shape recognition. The



**Figure 9.** The number of days the kilonova is detectable by ZTF. We use the same set of 1000 ideally sampled KN injections.

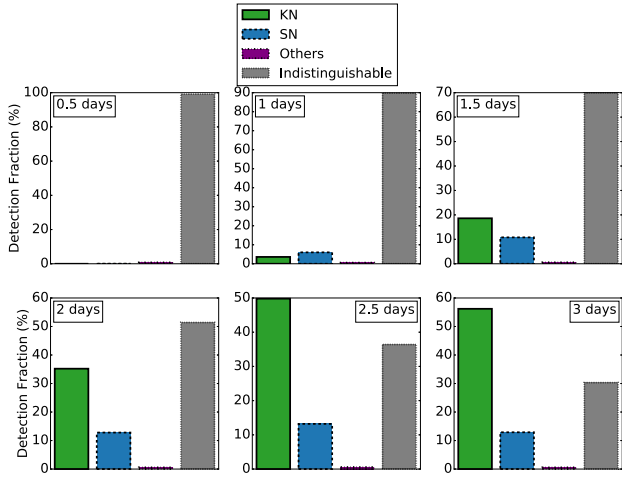
better classification efficiency with respect to the case of real ZTF ‘SN’-type objects seems to arise from the sampling cadence.

### 3.3.2 Well-sampled kilonovae

In this section, the kilonovae light curves considered are also ideally sampled (two observations per night taken in each consecutive night) with detections in two filters ( $r$  and  $g$  bands). Given their inherent faintness, the kilonovae are not visible at the ZTF sensitivity at large distances as the supernovae. More precisely, the injection set of kilonovae is visible for an average of 8.1 days. And the farther away a kilonova is, the less time it will be detectable by ZTF. A histogram showing the number of days the injected kilonovae are visible is in Fig. 9. As in the case of real ZTF objects, we consider for a kilonova visible for only  $N$  days that the preferred event after  $m$  days, with  $m > N$ , is the same as the preferred event at the end of the real  $N$  days of observation.

Fig. 10 presents the histograms of preferred events for the very first night of observations. As expected, the category ‘Indistinguishable’ is favoured when there are not enough observations (usually less than  $\sim 6$ ), but at the end of two nights, the classifier identifies it as a kilonova light curve with a high probability. From Fig. 10, two conclusions can be made. First, one can see that after only two days of observations, the ‘KN’ template starts to be preferred over the others. Secondly, one notices that after 3 days of observations, the classifier mainly chooses the ‘KN’ template, and the few failures consist of a preference for the ‘SN’ and ‘Indistinguishable’ templates. The false dismissal rate is less than 4 per 10 events. Fig. 11 shows the dependence of the classifier’s preferred events (at the end of 3 days of observation) as a function of  $X_{\text{lan}}$ , redshift,  $M_{\text{ej}}$ , and  $v_{\text{ej}}$ .

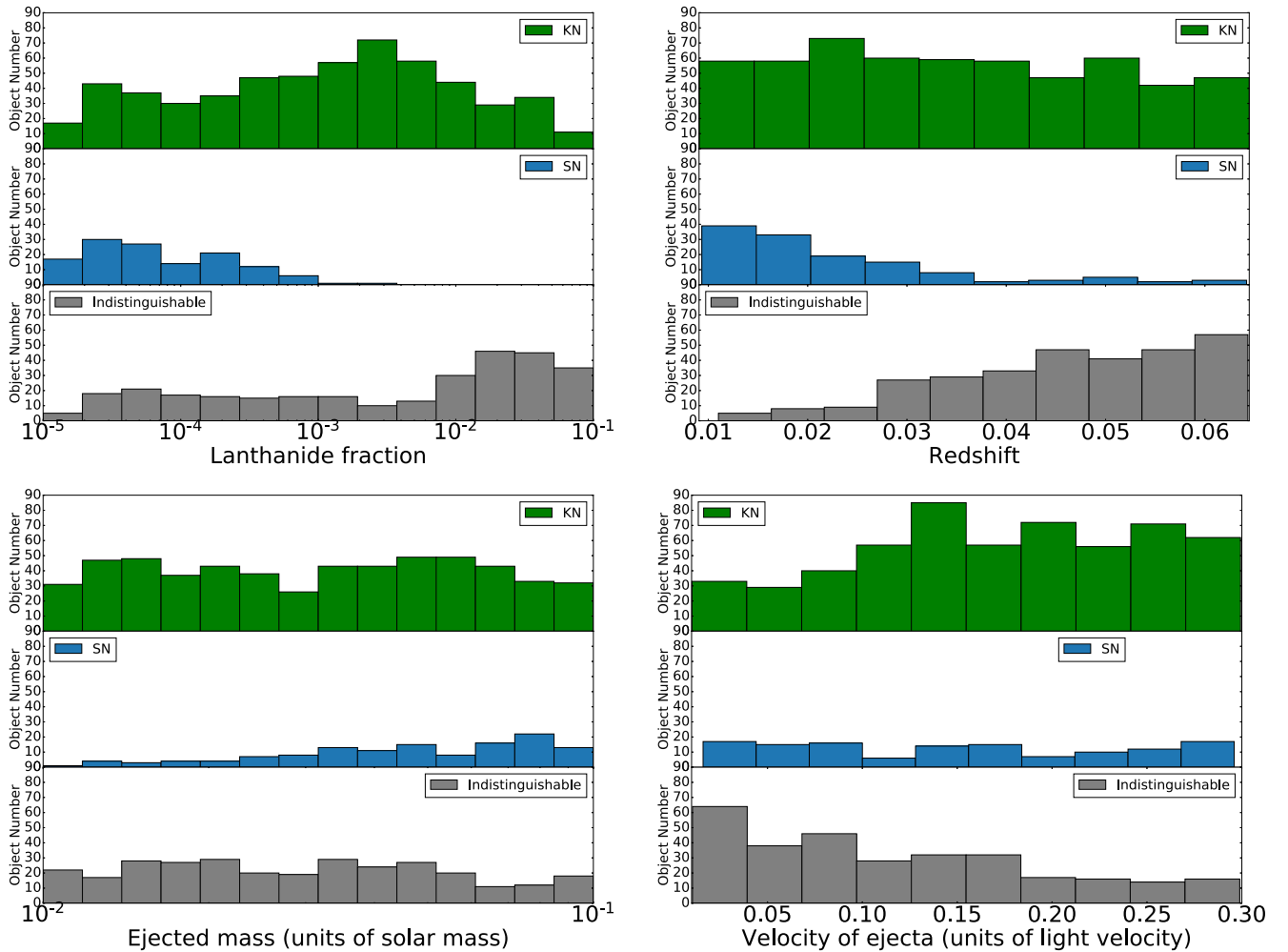
Fig. 11 suggests that a low velocity for the ejecta has the effect of preferring ‘Indistinguishable’ by the classifier. Likewise, the more mass ejected leads to a higher preference for the ‘SN’ templates; this is because they become brighter and therefore more consistent in appearance with supernovae. It also reveals that the lanthanide fraction and the redshift have a non-negligible impact on the light curve; we see that the smaller  $X_{\text{lan}}$ , the more probable is the ‘SN’ template, and the higher  $X_{\text{lan}}$ , the more probable is ‘Indistinguishable’. The lanthanide fraction is responsible for the reddening of the light curve. As the lanthanide fraction decreases, the transient becomes more blue and therefore compatible with a



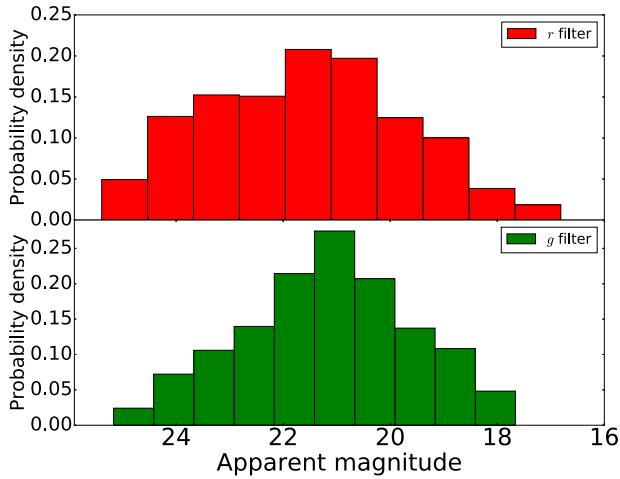
**Figure 10.** Preferred event fraction histogram given different amounts of observation time: from 0.5 day (top-left corner) up to 3 days (bottom-right corner). The input is represented by the set of 1000 kilonova injections.

supernova light curve (which are in general more grey). In addition, a large lanthanide fraction leads to red light curves, where the  $g$ -band observations are upper limits instead of detections and so the ‘Indistinguishable’ template becomes the favoured one, as there is no colour information. At small redshifts, kilonovae can be misclassified as SNe (intrinsically brighter) at higher distance with respect to the injected one. On the other hand, a high value of the redshift denotes a small signal-to-noise ratio, and thus the classifier is also more likely to prefer the ‘Indistinguishable’ template.

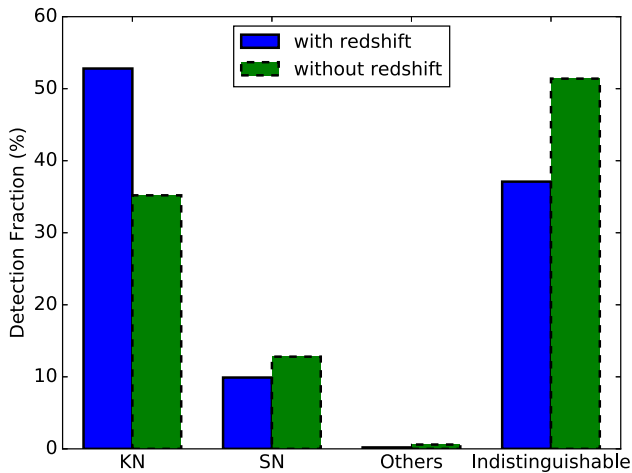
In order to assess the utility of our method, understanding the brightness of the kilonovae once they are detected for the first time by the algorithm is of great significance. To this end, we pick out only those light curves that were classified as ‘KN’ after the last observation. There are 757 such injections. For each of these light curves, we looked for the first time  $t_1$  when our tool identifies it as a ‘KN’, and then we determine the magnitude of the kilonova at this particular time  $t_1$ . Whereas all of these light curves contain observations in the  $r$  passband at  $t_1$ , only 553 of them are still detectable in the  $g$  passband at  $t_1$ . This is because in general the fade in  $g$  is faster than the fade in  $r$ , so usually that is the filter being measured last. In Fig. 12, we show the normalized histogram of the apparent magnitude with both filters at the time  $t_1$ . We can infer



**Figure 11.** Classification result dependence on the varied parameters for the 1000 KN injections. On the top row, a histogram of classifier favoured events after 3 days of observation as a function of  $X_{\text{lan}}$  (on the left) and redshift (on the right). On the bottom row is the same as a function of  $M_{\text{ej}}$  (on the left) and  $v_{\text{ej}}$  (on the right).



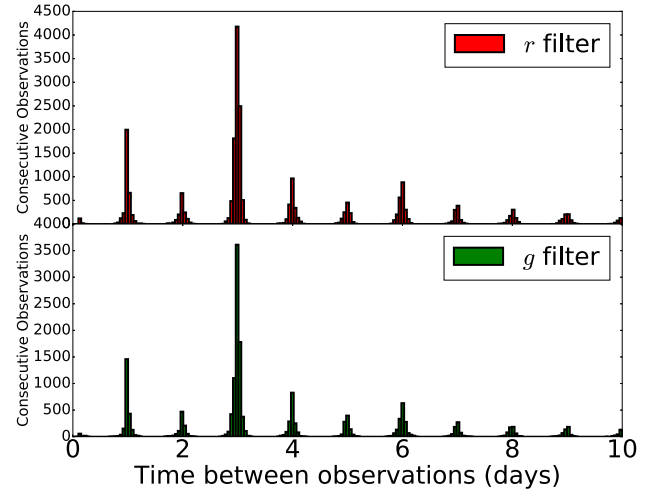
**Figure 12.** The probability density function for the apparent magnitude (on top for the  $r$  filter and on bottom for the  $g$  filter) at  $t_1$ , the first time the kilonova was identified as being ‘KN’. It was considered a sample of 757 (553) light curves for the  $r$  ( $g$ ) filter.



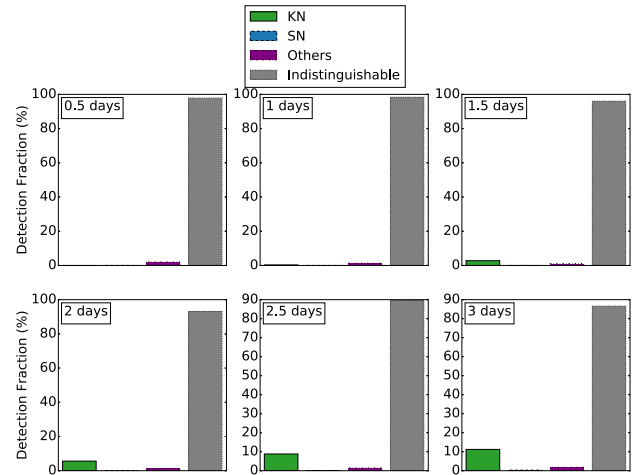
**Figure 13.** Classification results after two nights of observations with (blue) and without (green) the knowledge of the cosmological redshift. The input is represented by the same 1000 ‘KN’ injection set discussed inside this section.

from this figure that in the majority of cases, our tool identifies the kilonovae early enough for follow-up by telescopes that can reach  $\sim 21$  mag, i.e. reasonably sensitive telescopes.

In the previous paragraphs, we presented the results of our method when the input consists only of photometric data. However, as any classifier based on machine learning, *astrophysic* provides more precise classifications if additional information is available. One example of such additional information might be the redshift of the candidate. In Fig. 13, this is illustrated by means of a histogram for the quantitative improvement in the classification results when the redshift is taken into account compared to the situation where this information is considered unknown. For this study, we use the same set of well-sampled kilonova injections presented at the beginning of this section. One can observe that at the end of two nights of observations, the success rate in the recovery of KN templates is improved by more than 10 % if the cosmological redshift is given.



**Figure 14.** Histogram of time spacing between two consecutive observations in each filter. All real ZTF objects have been used for this study. The  $r$ -band results are on top and the  $g$ -band results are on bottom.



**Figure 15.** Preferred event fraction histogram given different amounts of observation time: from 0.5 day (top-left corner) up to 3 days (bottom-right corner). The input is represented by the set of 1000 kilonova injections sampled similarly to the ZTF cadence.

### 3.3.3 Realistically sampled kilonovae

While the case of dedicated ToOs was treated in the previous section by considering an ideal sampling of two observations per night and in each of the two filters  $r$  and  $g$  bands, we are now interested in what is the efficiency of the classification in the case of a serendipitous observation. To this end, the spacing in time of two consecutive observations was studied for the case of real ZTF objects. Indeed in Fig. 14, one can see the time difference between two consecutive observations in each filter for the entire set of real ZTF objects, both ‘SN’ and ‘Others’ types. One can easily observe a peak corresponding to a time spacing of 3 days. This is why we simulated a second set of 1000 kilonovae, with a sampling rate of one observation every three nights in each filter. Fig. 15 shows the results of the classification for the first 3 days of observations for this new set of injections. One can see that the classifications are much less efficient than in the case of the ideal sampling, which allows us



to conclude that dedicated ToOs versus serendipitous observations make a big difference.

#### 4 CONCLUSION

In this study, we present a method to classify transients starting from photometric observations and the cosmological redshift. The method is based on the use of an open-source classifier *astorapid*. By running this tool on input data represented by the observational light curves and combining the output results by source class, we propose a way to distinguish between four main classes: ‘KN’, ‘SN’, ‘Others’, and ‘Indistinguishable’. The performance of this classifier and class system has been tested on both real ZTF objects and simulated light curves. The case of real transients from the public ZTF alert stream emphasizes the necessity of around 10 observations at the ZTF cadence, provided that the information from both *r* and *g* passbands is used. Concerning the well-sampled light curves, it has been shown that the identification of SNe necessitates only a few observations, while for the recognition of KNe, a few nights of photometry in multiple passbands is required (with significantly worse results in the case of single passband observations). Finally, it has been shown that for kilonovae sampled at a cadence similar to that of ZTF, the efficiency of the classifier decreases significantly.

This study opens prospects for future work, including the dominant question in the community concerning how to determine the best observing strategy to adopt in the case of a GW alert. In order to address this question, several issues should be considered. First of all, an evaluation of the candidate recognition dependence on the observing cadence needs to happen. In addition, a performance comparison between imaging in two filters and spending twice as long in a single filter needs to happen (with of course the possibility of even more filters being considered). Also, the possibility of having three or more filters cannot be discredited. Once the entire parameter space is explored, an optimal observing strategy could be found. The current version of *astorapid* does not have the current capability to address these items. To do so, we will need to retrain *astorapid* on single passband light curves, as well as light curves observed in more than three filters.

In the future, we intend to improve *astorapid* to account for ‘missing’ observations where only upper limits are available. This should be useful in cases of particularly red or blue transients, where likely the colour information is even *more* apparent than in the case where there are detections. Another improvement to *astorapid* might consist of the introduction in the training set of new interlopers such as M-dwarf. In addition, we plan to incorporate these techniques into some of the ongoing follow-up infrastructure, such as the GROWTH target of opportunity marshal (Coughlin et al. 2019c) and the GRANDMA iCARE pipeline (Antier et al. 2020).

#### ACKNOWLEDGEMENTS

MC is supported by the David and Ellen Lee Prize Postdoctoral Fellowship at the California Institute of Technology. The authors thank the Observatoire de la Côte d’Azur for support.

#### DATA AVAILABILITY

The data underlying this article are derived from public code found here: <https://github.com/mcoughlin/gwemlightcurves>. The simulations resulting will be shared on reasonable request to the corresponding author.

#### REFERENCES

- Abbott B. P. et al., 2017a, *Phys. Rev. Lett.*, 119, 161101  
 Abbott B. P. et al., 2017b, *Nature*, 551, 85  
 Abbott B. P. et al., 2017c, *ApJ*, 848, L13  
 Abbott B. P. et al., 2017d, *ApJ*, 848, L12  
 Andreoni I. et al., 2020, *ApJ*, 890, 131  
 Antier S. et al., 2020, *MNRAS*, 492, 3904  
 Ascenzi S. et al., 2019, *MNRAS*, 486, 672  
 Barbary K. et al., 2016, Astrophysics Source Code Library, record ascl:1611.017  
 Bauswein A., Baumgarte T. W., Janka H.-T., 2013, *Phys. Rev. Lett.*, 111, 131101  
 Bauswein A. et al., 2017, *ApJ*, 850, L34  
 Bellm E. C. et al., 2018, *PASP*, 131, 018002  
 Berger E. et al., 2009, *ApJ*, 699, 1850  
 Berry C. P. L. et al., 2015, *ApJ*, 804, 114  
 Bloemen S. et al., 2016, in Hall H. J., Gilmozzi R., Marshall H. K., eds, Proc. SPIE Conf. Ser. Vol. 9906, Ground-Based and Airborne Telescopes VI. SPIE, Bellingham, p. 2118  
 Cornish N. J., Littenberg T. B., 2015, *Class. Quantum Gravity*, 32, 135012  
 Coughlin M., Dietrich T., Kawaguchi K., Smartt S., Stubbs C., Ujevic M., 2017, *ApJ*, 849, 12  
 Coughlin M. W. et al., 2018a, *MNRAS*, 480, 3871  
 Coughlin M. W. et al., 2018b, *MNRAS*, 478, 692  
 Coughlin M. W. et al., 2019a, *PASP*, 131, 048001  
 Coughlin M. W., Dietrich T., Heinzel J., Khetan N., Antier S., Christensen N., Coulter D. A., Foley R. J., 2019b, *Phys. Rev. Research*, 2, 022006  
 Coughlin M. W., Dietrich T., Margalit B., Metzger B. D., 2019c, *MNRAS*, 489, L91  
 Coughlin M. W. et al., 2019d, *ApJ*, 885, L19  
 Coulter D. A. et al., 2017, *Science*, 358, 1556  
 Dietrich T., Ujevic M., 2017, *Class. Quantum Gravity*, 34, 105014  
 Drout M. R. et al., 2017, *Science*, 358, 1570  
 Ducoin J.-G., Corre D., Leroy N., Le Floch E., 2020, *MNRAS*, 492, 4768  
 Essick R., Vitale S., Katsavounidis E., Vedovato G., Klimentenko S., 2015, *ApJ*, 800, 81  
 Fairhurst S., 2009, *New J. Phys.*, 11, 123006  
 Fairhurst S., 2011, *Class. Quantum Gravity*, 28, 105021  
 Foley R. J. et al., 2013, *ApJ*, 767, 57  
 Goldstein A. et al., 2017, *ApJ*, 848, L14  
 Graham M. J. et al., 2019, *PASP*, 131, 078001  
 Grover K., Fairhurst S., Farr B. F., Mandel I., Rodriguez C., Sidery T., Vecchio A., 2014, *Phys. Rev. D*, 89, 042004  
 Guy J. et al., 2007, *A&A*, 466, 11  
 Ishida E. E. O. et al., 2019, *MNRAS*, 483, 2  
 Ivezić Z. et al., 2019, *ApJ*, 873, 111  
 Kasen D., Metzger B., Barnes J., Quataert E., Ramirez-Ruiz E., 2017, *Nature*, 551, 80  
 Kessler R. et al., 2009, *PASP*, 121, 1028  
 Kessler R. et al., 2019, *PASP*, 131, 094501  
 Klimentenko S. et al., 2016, *Phys. Rev. D*, 93, 042004  
 Lattimer J. M., Schramm D. N., 1974, *ApJ*, 192, L145  
 LIGO Scientific Collaboration, Virgo Collaboration, 2019, GRB Coordinates Netw., 24168, 1  
 Li L.-X., Paczynski B., 1998, *ApJ*, 507, L59  
 Lunnan R. et al., 2017, *ApJ*, 836, 60  
 Mészáros P., Rees M. J., 1998, *ApJ*, 502, L105  
 Metzger B. D., Berger E., 2012, *ApJ*, 746, 48  
 Metzger B. D. et al., 2010, *MNRAS*, 406, 2650  
 Morgan J. S., Kaiser N., Moreau V., Anderson D., Burgett W., 2012, in Larry M. S., Roberto G., Helen J. H., eds, Proc. SPIE Conf. Ser. Vol. 8444, Ground-Based and Airborne Telescopes IV. SPIE, Bellingham, p. 0H  
 Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019, *PASP*, 131, 118002  
 Pian E. et al., 2017, *Nature*, 551, 67

Piran T., Nakar E., Rosswog S., 2013, *MNRAS*, 430, 2121  
 Quimby R., Aldering G., Wheeler J., H'offlich P., Akerlof C., Rykoff a., 2008, *ApJ*, 668, L99  
 Radice D., Perego A., Zappa F., Bernuzzi S., 2018, *ApJ*, 852, L29  
 Rees M. J., 1988, *Nature*, 333, 523  
 Ren J., Christlieb N., Zhao G., 2012, *Res. Astron. Astrophys.*, 12, 1637  
 Röser C., Meyer R., Guidi G. M., Viceré A., Christensen N., 2007, *Class. Quantum Gravity*, 24, S607  
 Savchenko V. et al., 2017, *ApJ*, 848, L15  
 Shen K. J., Kasen D., Weinberg N. N., Bildsten L., Scannapieco E., 2010, *ApJ*, 715, 767  
 Sidery T. et al., 2014, *Phys. Rev. D*, 89, 084060

Silverman J. M., et al., 2012, *MNRAS*, 425, 1789  
 Singer L. P. et al., 2014, *ApJ*, 795, 105  
 Smartt S. et al., 2017, *Nature*, 551, 75  
 Tanvir N., Levan A., Fruchter A., Hjorth J., Wiersema K., Tunnicliffe R., de Ugarte Postigo A., 2013, *Nature*, 500, 547  
 Tonry J. L. et al., 2018, *PASP*, 130, 064505  
 Troja E. et al., 2018, *Nat. Commun.*, 9, 4089  
 Troja E. et al., 2018, *MNRAS*, 478, L18  
 Watson D. et al., 2019, *Nature*, 574, 497  
 Wen L., Chen Y., 2010, *Phys. Rev. D*, 81, 082001  
 Wijers R. A. M. J., Rees M. J., Mészáros P., 1997, *MNRAS*, 288, L51

## APPENDIX: PENALTY FACTOR AND CHOICE OF THRESHOLD

When collapsing the probabilities of the 14 initial *astorapid* templates, one could naively sum the corresponding probabilities, i.e.

$$P(\text{SN}) = P(\text{SNIa-norm}) + P(\text{SNIbc}) + P(\text{SNIc}) + P(\text{SNIa-91bg}) + P(\text{SNIa-x}) + P(\text{point-Ia}) + P(\text{SLSN-I}) + P(\text{PISN})$$

$$P(\text{Others}) = P(\text{ILOT}) + P(\text{CART}) + P(\text{TDE}) + P(\text{AGN})$$

$$P(\text{KN}) = P(\text{Kilonova})$$

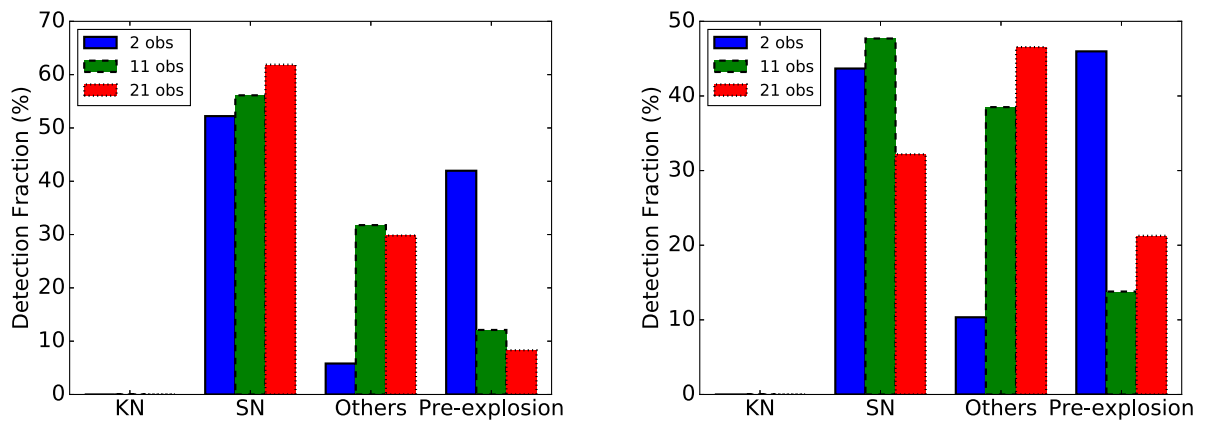
$$P(\text{Pre-explosion}) = P(\text{Pre-explosion}).$$

Given the abundance of SN classes, ‘SN’ would almost always be selected as the preferred class.

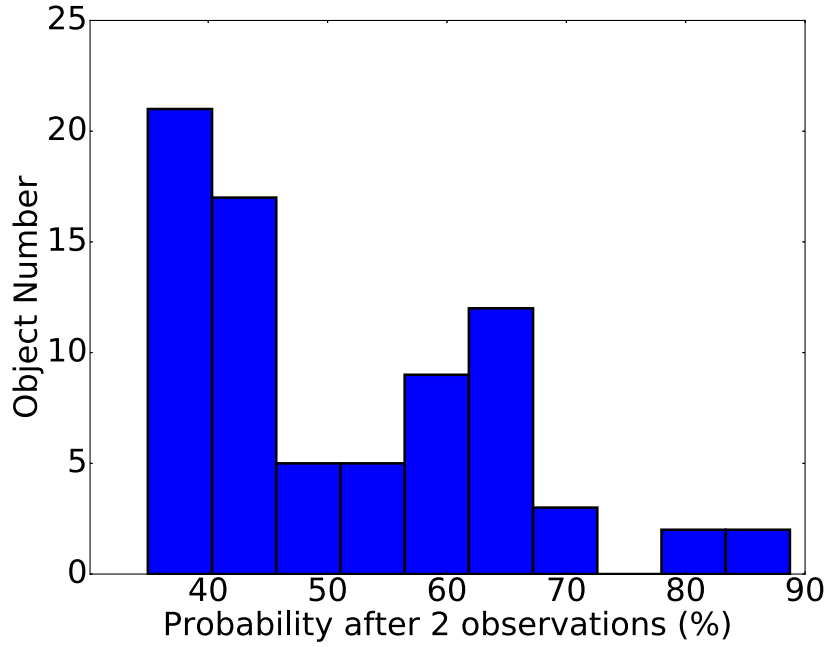
We illustrate this behaviour on real ZTF light curves, i.e. the same data set we used in Section 3.1. In Fig. A1, we plot the detection fraction of the preferred event after different numbers of observations. From this figure, several conclusions can be drawn. First of all, one can see, especially in the case where ‘Others’ light curves are input, the ‘SN’ class is often incorrectly chosen. This behaviour is visible for early observations especially and improves somewhat at later times. This motivates our decision to penalize the ‘SN’ class by a factor  $(1 - e^{-k_{\text{obs}}^{\text{th}}/\beta})$ , where  $k_{\text{obs}}^{\text{th}}$  stands for the  $(k + 1)$ th observation. Secondly, at early times, too many events are misclassified as ‘Others’ and ‘SN’, when a non-determinative choice, i.e. a preference for the ‘Pre-explosion’ template, is physically reasonable. This supports our decision to replace ‘Pre-explosion’ with ‘Indistinguishable’ as well as the introduction of a probability threshold in order for a transient to be classified in the set {‘Indistinguishable’, ‘SN’, and ‘Others’}.

In Fig. A2, we illustrate the distribution of *astorapid* preferred event probabilities after two observations, when the input is a real ZTF ‘Others’-type object, and the preferred event at the end of two observations is in {‘KN’, ‘SN’}. Fig. A2 therefore shows the probability of the ‘wrong’ preferred class when there is misclassification at early times. As a consequence of these results, we choose a threshold equal to 40 % below which the preferred event is always ‘Indistinguishable’.

Once the threshold is fixed, we are looking for a value of  $\beta$  based on our data set results. Given the expression of the penalty factor we imposed,  $(1 - e^{-k_{\text{obs}}^{\text{th}}/\beta})$ , it is worth mentioning the  $\beta$ -dependence of our classifications. A too small  $\beta$  means a penalty factor close to 1, so basically the classifications will be very similar to our initial results. On the other hand, a too large  $\beta$  means a penalty factor close to 0, which will have as a consequence the preference of the other classes than ‘SN’. We can anticipate from this discussion that a near-optimal  $\beta$  will be obtained as a trade-off between these two regimes. To assess this, we analyse the impact of the penalty factor on the same set



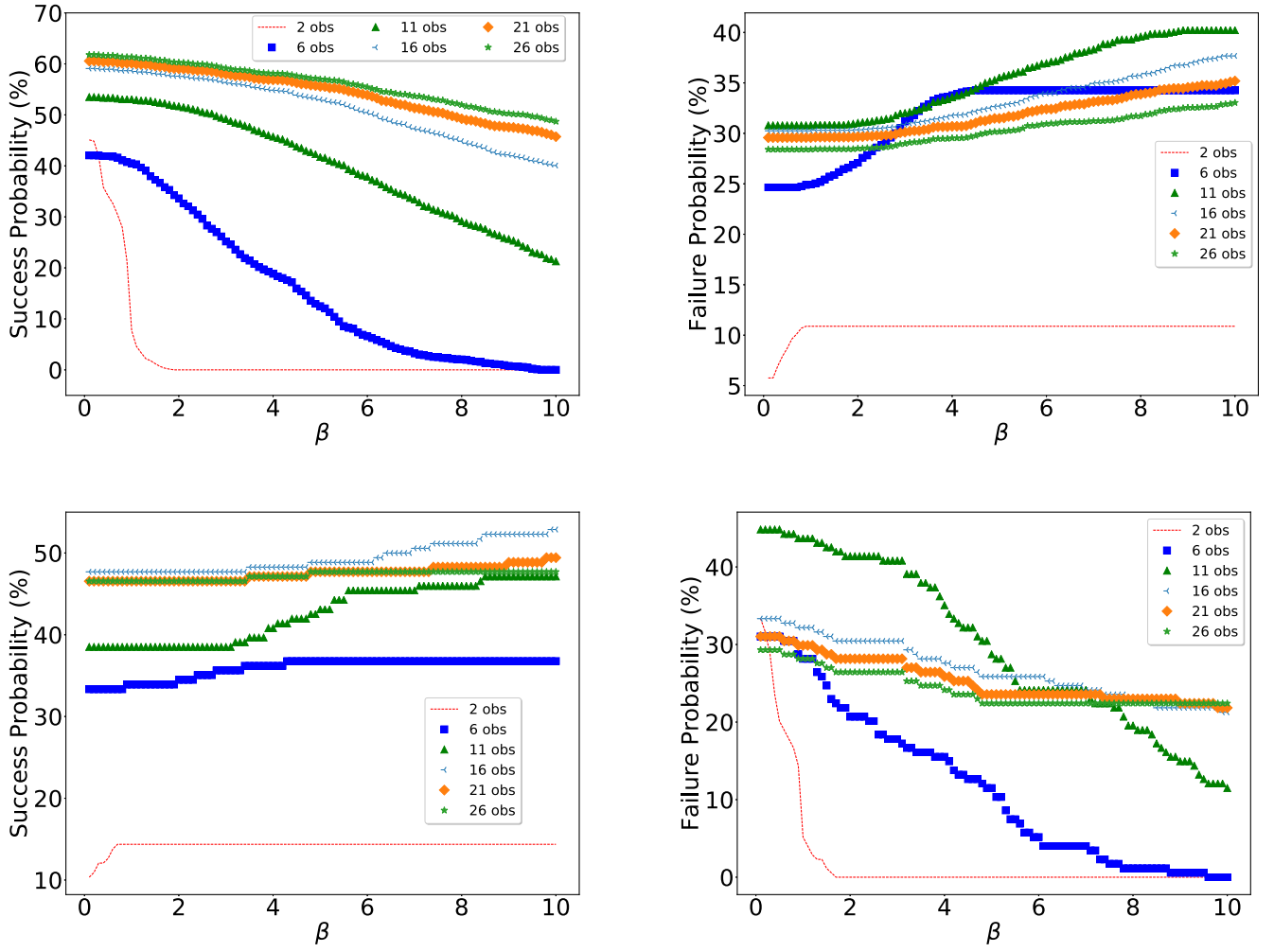
**Figure A1.** The preferred event after 2, 11, and 21 observations, respectively. The input is represented on the left by the set of 2049 ZTF real sources identified as ‘SN’ type and on the right by the set of 174 ZTF real sources identified as the ‘Others’ type.



**Figure A2.** Histogram of the preferred event probabilities after the two observations. The input is represented by those real ‘Others’ light curves that are misclassified as ‘SN’ or ‘KN’ after two observations.

of real ZTF objects as presented in Section 3.1. For the ‘SN’-type objects, we define the failure probability as being the proportion of ‘SN’-type objects classified as ‘Others’ or ‘KN’. Analogously for the ‘Others’-type objects, we define the failure probability as the proportion of ‘Others’-type objects classified as ‘SN’ or ‘KN’. It should be clear that these success/failure probabilities are a function of the number of observations.

In Fig. A3, one can see that except for the very earliest times ( $\sim 2$  observations), all the other curves contain  $\beta = 4$ , indicating that it is near-optimal.



**Figure A3.** Success and failure probability for ‘SN’-type objects (on top) and ‘Others’-type objects (on bottom). It turns out from these plots that  $\beta = 4$  is a reasonable choice.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.